

# WalkieLokie: Relative Positioning for Augmented Reality Using a Dummy Acoustic Speaker

Wenchao Huang, Yan Xiong, Xiang-Yang Li, Yiqing Hu, Xufei Mao, Panlong Yang



**Abstract**—We propose and implement a novel relative positioning system, WalkieLokie, to enable more kinds of Augmented Reality applications, e.g., virtual shopping guide, virtual business card sharing. WalkieLokie calculates the distance and direction between an inquiring user and the corresponding target. It only requires a dummy speaker binding to the target and broadcasting inaudible acoustic signals. Then the user walking around can obtain the position using a smart device. The key insight is that when a user walks, the distance between the smart device and the speaker changes; and the pattern of *displacement* (variance of distance) corresponds to the relative position. We use a second-order phase locked loop to track the displacement and further estimate the position. To enhance the accuracy and robustness of our strategy, we propose a synchronization mechanism to synthesize all estimation results from different timeslots. We show that the mean error of ranging and direction estimation is 0.63m and 2.46 degrees respectively, which is accurate even in case of virtual business card sharing. Furthermore, in the shopping mall where the environment is quite severe, we still achieve high accuracy of positioning *one dummy* speaker, and the mean position error is 1.28m.

## 1 INTRODUCTION

With the rapid development of smart phones and wearable devices, attractive Augmented Reality (AR) apps have been developed, e.g., Sky Map, Wikitude, Augmented Car Finder. One of AR's key features is to display useful information about a person's surroundings, which relies on location information. For example, Wikitude uses GPS and inertial sensors to provide interactive information about objects that are seen through the camera of smart devices.

In this paper, we explore localization techniques to enable more kinds of AR applications on smart devices. For instance, a person walks in a large shopping mall and a virtual shopping guide recommends the surrounding goods that are new arrivals or on sale; or shares her/his virtual business card with people walking around in

a party. Such applications require the knowledge of *relative position* between targets (e.g., goods, person) and inquiring users.

However, current localization systems cannot be directly applied to *relative positioning* satisfactorily due to various limitations. For instance, these systems can only be used in some places with specified infrastructure being deployed, or require feature-rich hardware serving as target. More specifically, GPS can calculate the location of outdoor users, but is unavailable in indoor environments. Pure WiFi-based indoor localization can achieve 3~4 meter accuracy in absolute positioning and there always exists large errors (e.g., 6~8m) [6]. So the errors are much greater than 4 meters when inferring *relative position* from *absolute positions* of the smartphone and the target. Other indoor localization schemes [7], [19] are accurate enough (e.g., sub-meter accuracy), but require special-purpose infrastructure or hardware. There are schemes calculating relative direction (e.g., Swadloon [3]) and distance (e.g., BeepBeep [12]). However, Swadloon requires unusual behavior of querying user (i.e., phone-shaking movement) before getting the direction of a target. More importantly, Swadloon cannot obtain distance from the target. Though BeepBeep can be added for calculating distance, it requires that the target has rich functions, such as broadcasting and receiving acoustic signals, communication for exchanging data and computation for processing data. It is feasible when the target uses a smartphone which has all these functions, but other applications, such as shopping guide, may prefer a cheaper target device with much fewer functions.

We propose and implement WalkieLokie, which calculates the *relative position* from a user with a smart device to a target for Augmented Reality. WalkieLokie does not need any infrastructure being deployed, and the application of WalkieLokie is not limited by places. The only requirement of WalkieLokie is that the target is attached with a dummy speaker for broadcasting audio, which can be received by the smart device and then processed to directly infer the relative position. The dummy speaker merely broadcast audio without requirement of any other features, e.g., audio recording, communication or computation. Hence, they are widely used and some of them are cheap and simple, such as speaker embedded in user's smart devices, or even

Wenchao Huang, Yan Xiong, Yiqing Hu are with School of Computer Science and Technology, University of Science and Technology of China. Email: {huangwc, yxiong, mfy}@ustc.edu.cn, huyiqing@mail.ustc.edu.cn.

Xiang-Yang Li is with Department of Computer Science and Technology and TNLIST, Tsinghua University, and Department of Computer Science, Illinois Institute of Technology. Email: xli@cs.iit.edu.

Xufei Mao is with Department of Software Engineering, and TNLIST, Tsinghua University. Email: xufei.mao@gmail.com.

Panlong Yang is with Institute of Communication Engineering, PLAUST. Email: panlongyang@gmail.com.

loudspeaker originally for sales promotion in a shopping mall. Moreover, the broadcast audio is inaudible that the loudspeaker, which used to be a noisy tool for sales promotion, can now be “silent” for the same job by “broadcasting” its relative position.

Our work is based on the observation that when a user walks, the distance between the object and the user changes; and the pattern of *displacement* (variance of distance) relates to the relative position. In other words, by letting a device receive and analyze the signal (audio signal mixed with non-audible signal) broadcast by a target (speaker), we are able to track the displacement and further compute the relative position accurately and efficiently, i.e., finishing both ranging and direction estimation at the same time.

However, we have to solve a number of issues in our scheme. First, since the displacement is relatively small, the practical challenge is how to obtain the position precisely when a user walks for only very few steps. Second, when the user is far from the speaker, the measured displacement is prone to be influenced by noises and the difference of displacement becomes more indistinguishable such that even a tiny error in the measurement could cause large errors in positioning. Hence, both accurate displacement measurement scheme and robust positioning strategy are needed.

In order to get an accurate displacement measurement scheme, we track the phase of the signal (corresponding to the displacement) utilizing the second-order Phase Locked Loop (PLL), which could avoid jitters and has high accuracy when the signal is weak. Hence, the distance and direction could be computed accurately when a user is close to a speaker. Next, as we have mentioned above, the estimated position may have a bigger error when a user is far from a speaker. In this case, we adopt two strategies to further improve the accuracy and robustness. One is to utilize the measurement results when the user is close to the speaker if available. Otherwise, we synthesize all the estimations (longer path a user passed) by the synchronizing scheme. The main idea is based on the following observation: Since the distance can be obtained according to the difference between the sent time and received time of a signal, where the receiving time is directly computed but the sent time is unknown for the receiver, we add a periodical pulse into the audio to get the sent time in a novel way. Specifically, when a good estimation is obtained, the distance along with the sent time of the pulse is calculated. Hence, the sent time of the later periodical pulses is predicted, which can infer the distance according to sent time and receiving time of the pulses.

WalkieLokie also addresses a number of practical issues based on the main solution:

**The user frequently turns the walking direction:** we provide enhanced algorithm that gathers all the pieces of small linear segments at different directions to obtain the position.

**Multipath effects on pulse detection:** WalkieLokie de-

fects arrival time of all the pulses, including the pulse directly from sender and also reflected ones. Then it eliminates the false pulses by leveraging the result of PLL.

**Non-Line-of-Sight (NLoS):** WalkieLokie uses historical position results, and infers the current position by additionally using inertial sensors; once the smart device is within the coverage of the signal, WalkieLokie updates accurate position by synchronization.

**Device diversity:** The main problem is serious clock drift of normal dummy speaker, otherwise the receiver obtains wrong receiving time of periodical pulses and further wrong distance in synchronization. We leverage the result of PLL, and calibrate the clock precisely in case that the receiver is static for only a few seconds.

**Device Orientation:** Different orientation of the speaker or receiver affects the quality of the received signal. We find that the quality mainly affects the result of PLL and further the displacement. More specifically, when the signal quality is poor for certain orientation, the tracked displacement becomes smaller than the real displacement. To enhance the accuracy, we make calibrations on tracked displacement based on our measurements.

**Conflicts of multiple signals:** In WalkieLokie, the periodical pulses in synchronization possess bandwidth, which limits the number of co-existing signals. We carefully design the pulse that possesses narrow bandwidth, and also show the way of supporting more number of co-existing speakers.

**Noisy environment:** WalkieLokie uses Band Pass Filter to eliminate the noises and it works well in the noisy shopping mall.

We implement WalkieLokie and evaluate the performance of all the components separately with several types of cases and then the performance of WalkieLokie. **a).** For the case when a user is within 8 meters away from a speaker, the mean error of ranging and direction estimation is  $0.63m$  and  $2.45^\circ$ . It shows considerable accuracy when a user shares virtual business card with surroundings.

**b).** When the user is within  $20m$  and uses synchronization for positioning, the ranging and direction estimation errors are less than  $0.32m$  and  $2.81^\circ$  at the percentage of 80% respectively. Note that the results in this case only infer the accuracy of the subcomponent (synchronization), instead of the total accuracy of WalkieLokie.

**c).** We combine all the work together and evaluate WalkieLokie in a severe environment, i.e., the noisy shopping mall. We conduct the experiment in 2 cases:

- **c1).** *Relative positioning of one speaker;*
- **c2).** *Absolute positioning using multiple speakers (i.e., ordinary indoor localization).*

We put 5 dummy speakers in a  $600m^2$  area, and the positions of speakers are limited to be deployed, (just at the side of aisle, instead of the position on the ceiling). Even in this case, by using only *one* of these speakers, WalkieLokie also achieves the mean error of  $1.28m$  for

the *relative* position. Hence, a user knows the accurate relative position of a virtual shopping guide attached with a dummy speaker.

To explore the possibility of *absolute* positioning, we also use all the 5 speakers as anchors for localization and the mean error is  $0.89m$ , where each position is covered by the signal of less than 2 speakers on average. Since normal anchor-based systems only obtain distance or direction but cannot calculate both metrics, they require at least 3 anchors for trilateration. Hence it shows another advantage of WalkieLokie that it is robust when the anchors are sparse in deployment.

The rest of the paper is organized as follows. We first present the overview of WalkieLokie in Section 2. We propose the position estimation based on displacement tracking in Section 3 and displacement tracking method in Section 4. We give the details of the synchronization in Section 5. We report our extensive experimental results in Section 6. We review some related work in Section 7. We conclude the paper in Section 8.

## 2 OVERVIEW

### 2.1 Problem Description

WalkieLokie calculates *relative position* between a user with a smart device and a target attached with a dummy speaker, where the relative position can decompose into distance and direction from the smart device to the dummy speaker. The dummy speaker merely broadcasts inaudible audio without the requirement of any other features. The smart device has a microphone and inertial sensors (*i.e.*, compass, accelerometer, gyroscope), which are common components in almost all smart devices.

### 2.2 Intuitive Solution

The key insight of our paper is that when a user walks along a line, the pattern of displacements from the user to a dummy speaker is related to relative position directly.

We illustrate the intuitive solution on a simple case in Figure 1a, where a user walks and steps at  $O_1$ ,  $O_2$ , and  $O_3$ . Suppose the displacements  $d_1(=l_1 - l_2)$  and  $d_2(=l_2 - l_3)$  are measured beforehand and the user's stride length ( $|\overline{O_1O_2}|$ ) is given. Intuitively,  $d_1 \approx 0$  infers that  $O_1$  and  $O_2$  are close to  $H$  where  $\overline{AH} \perp \overline{O_1O_2}$  and  $d_2 < 0$ , which tells us that the speaker is at the back of the walking user, hence infers the coarse-grained direction. Another observation is that when the distance  $|\overline{AH}|$  increases, the value of  $|d_2 - d_1|$  decreases, which infers the coarse-grained distance as well. So, the relative position between  $O_1$  and  $A$  can be estimated.

### 2.3 Main Technical Issues

From the above example, the following main technical issues required to be solved:

**Formal solution of relative positioning (Section 3).** Given the real-time relative displacement, we need to

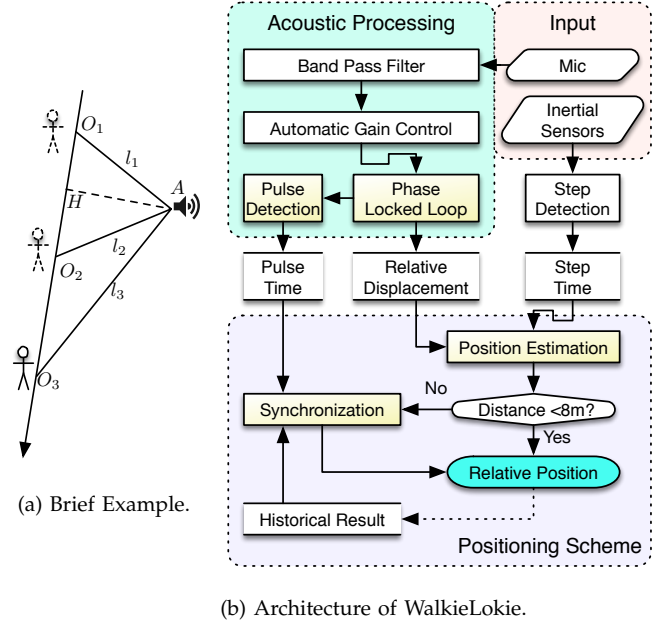


Fig. 1: Example of relations between displacement and relative positions, and architecture of WalkieLokie.

calculate the precise relative position, instead of coarse-grained one.

**Tracking relative displacement (Section 4).** The relative displacement needs to be tracked before relative positioning. Note that to the best of our knowledge, current approaches cannot directly obtain distances  $l_1$  without synchronization between the receiver and the dummy speaker. They require additional capabilities of the speaker, such as communication capability that exchanges synchronization information [12]. Instead, we calculate the displacement:  $d_1(=l_1 - l_2)$  by PLL.

**Extended solution when distance is longer (Section 5).** When  $|\overline{AH}|$  becomes much longer,  $|d_2 - d_1|$  is much smaller. Since there are errors on tracking the displacement  $d_1, d_2$ , the ideal case is that small change of distance corresponds to large value of  $|d_2 - d_1|$ , which results in high accuracy of calculated distance. However, when  $|\overline{AH}|$  becomes much larger, it is the opposite case that tiny error on measuring  $d_1$  or  $d_2$  will result in large error on calculating  $|\overline{AH}|$ . Hence, the accuracy of ranging declines when  $|\overline{AH}|$  becomes larger, and we need extended solution in this case. Note that the accuracy of direction finding is not much affected that we mainly propose the extended solution for ranging.

### 2.4 Architecture

To solve the technical issues, we divide WalkieLokie into 3 main components in Figure 1b: input of smart device, acoustic processing, and positioning scheme.

**Input:** The microphone and inertial sensors are used in WalkieLokie. The microphone records audio for acoustic processing. The inertial sensors mainly serve as a step counter, which records the time when the user steps on

the ground. When the user turns direction, the angle of user's rotation is also calculated by the gyroscope.

**Acoustic processing:** This component generates intermediate results preparing for the positioning scheme.

One result is *relative displacement*, which is tracked by analyzing the recorded audio (in Section 4). The audio firstly passes through the Band Pass Filter (BPF) that the signal at the specified frequency passes and other signals including human voice and noises are eliminated. Then, the filtered signal is processed by Automatic Gain Control (AGC), and then the amplitude of the signal is close to constant. The signal then passes through our carefully-designed Phase Locked Loop (PLL), and the phase of the signal, which is proportional to relative displacement, is tracked.

Another intermediate result provides additional information for the extended solution. More specifically, we encode periodical pulses in the sent signal, and the smart device detects the corresponding pulses to determine the *receiving time of the pulses* (in Section 5). The problem is that the pulse should take very little bandwidth, otherwise the number of concurrent speakers is much limited. We carefully encode the signal to solve this problem, and design the pulse detection algorithm to precisely determine receiving time of the pulses. Note that the pulse detection analyzes tracked phase from output of PLL, for we directly modulate the phase, rather than the raw audio, to encode the pulses in order to save bandwidth.

**Positioning scheme:** The scheme calculates position by receiving the intermediate results. The scheme firstly estimates position by using the relative displacements and user's step time. It leverages the intuitive solution in Section 2.2, which is formally illustrated in Section 3. Then, if the computed distance is very short ( $< 8m$ ), the calculated position is accurate enough and accepted as valid result. Otherwise, the calculated direction is accurate, but the calculated distance is inaccurate. In this case, the scheme invokes synchronization in Section 5 to compute the relative position. The synchronization uses the historical results of relative position to infer the distance. By additionally using the historical receiving time of the pulses, the sending time of the periodical pulses is then calculated. The accurate distance is then inferred from the detected receiving time and the predicted sending time of the current pulse.

### 3 ESTIMATION OF RELATIVE POSITION

In this section, we propose the method on distance and direction estimation from smart device to speaker, *i.e.*, the relative position. The intuition is that when the user walks, there is a unique pattern of displacement according to relative position. Hence, we use the displacements (calculated in section 4) to deduce user's positions.

#### 3.1 Positioning When User Walks along a Line

To estimate the distance, we first consider a simple scenario when a user starts walking from  $O_1$  and steps at

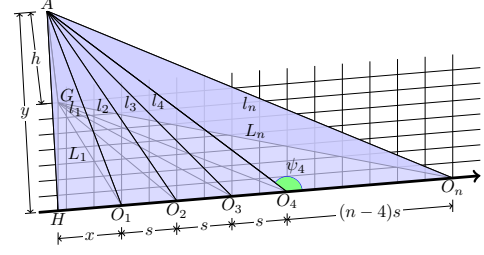


Fig. 2: Positioning when the user walks along a line.

$O_2, O_3, \dots, O_n$ , shown in Figure 2.  $A$  is the position of the speaker. Denote the height of the speaker relative to the smart device as  $h = |AG|$ . Assume the stride length is close to constant  $s = \overline{O_i O_{i+1}}$ . Both  $h$  and  $s$  are assumed to be given and used for distance and direction estimation. The other inputs are the displacements of all the steps, *i.e.*,  $d_i = l_i - l_{i+1}$  for the step  $\overline{O_i O_{i+1}}$ , which are calculated using PLL. Observe that the distance from the speaker to  $\overline{O_i O_{i+1}}$  is constant  $y = |AH|$ , where  $AH \perp \overline{O_1 O_n}$ . Hence, we first estimate  $x = |HO_1|$  and  $y$  from those inputs and then estimate the position at each step point  $O_i$  according to  $x$  and  $y$ .

Intuitively,  $x$  and  $y$  can be found by traversing the positions and using maximum likelihood estimation. Specifically, as  $|HO_i| = x + (i-1)s$ ,  $i = 1, 2, 3, \dots$ , denoting that

$$l'_i = \sqrt{y^2 + (x + (i-1)s)^2} \quad (1)$$

$$e_i = l'_i - l'_{i+1} - d_i \quad (2)$$

Then  $e_i = 0$  if  $d_i$  is accurate. Hence, for  $n$  displacements  $d_1, d_2, \dots, d_n$ ,  $x$  and  $y$  can be solved from above  $n$  equations by  $(x, y) = \arg \min_{x, y} \sum_{i=1}^n e_i^2$ . Here we use the Newton's Method [9] to reduce the computation overhead.

Observe that  $L_i = |\overline{GO_i}|$  and  $\psi'_i = \angle GO_i O_n$ , instead of  $l_i$  and  $\psi_i$ , are the horizontal distance and direction and used for positioning when  $x$  and  $y$  are estimated, we make the distance and direction results in the following equations:

$$\begin{cases} L_i = \sqrt{(x + (i-1)s)^2 + y^2 - h^2} \\ \cos \psi'_i = -\frac{x + (i-1)s}{L_i} \end{cases} \quad (3)$$

#### 3.2 Synthesizing When User Turns Direction

When a user turns direction while walking, we can always calculate the relative position as follows. Assume that the user starts from  $O_a$  and walks along the linear segment  $\overline{O_a O_b}, \overline{O_b O_c}, \overline{O_c O_d}, \overline{O_d O_e}$  in Figure 3. We use the calculated displacements in this case. We also use the step counter to estimate the linear length  $n_a s, n_b s, n_c s$ , where  $s$  is the stride length and  $n_a$  is the number of steps when user walks from  $O_a$  to  $O_b$ .

**Calculating angle of turning direction:** We calculate angle of user's rotation mainly by using gyroscope, when the user turns direction. Though Zee [15] can directly calculate walking direction, it is mainly affected





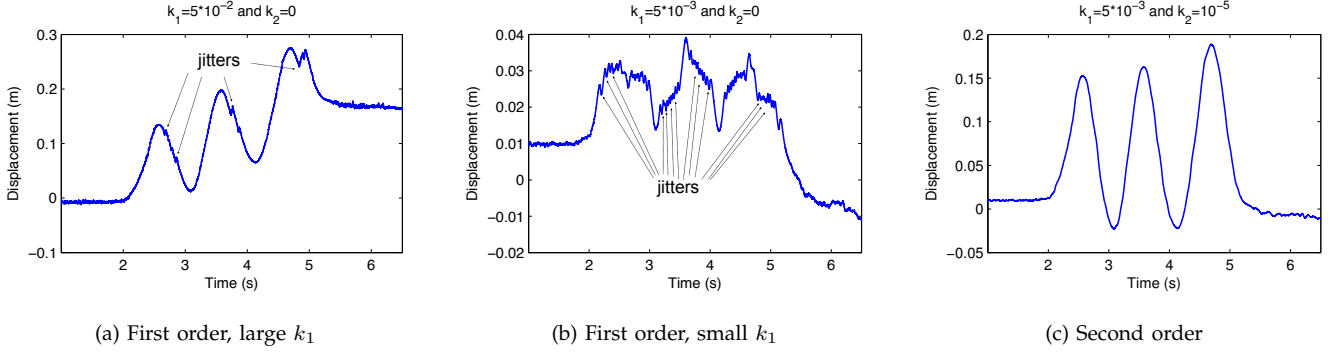


Fig. 4: Calculated displacement by PLL with different orders and parameters, when the signal is weak.

### 4.3 Preprocessing Received Signal

Before tracking phase  $\phi$  from  $r_1(t)$ , we have to preprocess the received signal. For the sent signal  $s_1(t)$ , the actual received signal  $r_{raw}(t)$  does not equal to  $r_1(t)$ . Its amplitude  $A(t)$  always changes and it is also mixed with noises  $\sigma(t)$ . We denote  $r_{raw}(t) = A(t) \cos(2\pi ft + \phi(t)) + \sigma(t)$ . Hence, we need to firstly eliminate  $A(t)$  and  $\sigma(t)$  before tracking the phase  $\phi(t)$ .

To eliminate the noise  $\sigma(t)$ , we let  $r_{raw}(t)$  pass through a Band Pass Filter (BPF). The processed signal  $r_{filter} \approx A(t) \cos(2\pi ft + \phi(t))$ .  $r_{filter}$  is then processed by Automatic Gain Control (AGC) [16]. After that,  $A(t)$  is removed and the signal can be seen as  $r_1(t)$  [3].

### 4.4 Tracking the Phase

To track phase for inferring displacement, we adopt the second-order Phase Locked Loop (PLL) to track the phase when the smart device moves, rather than the ordinary first-order PLL. PLL is a classical method in signal processing and can be regarded as a device that tracks the phase and frequency of a sinusoid. In our design, it is implemented purely by software due to the limited capabilities of smartphone platform.

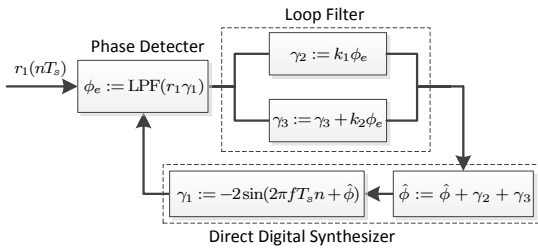


Fig. 5: Design of the Second-Order Phase Locked Loop.

We show our design of PLL in Figure 5. The PLL contains three main components: phase detector, loop filter and direct digital synthesizer (DDS). The phase detector detects the difference  $\phi_e = \phi - \hat{\phi}$ , where  $\hat{\phi}$  is the estimation of  $\phi$ . According to  $\phi_e$ , the loop filter analyzes and predicts the offset  $\gamma_2 + \gamma_3$  of  $\hat{\phi}$  for the next cycle of the loop, where the variance of  $\gamma_2, \gamma_3$  is affected by parameter  $k_1$  and  $k_2$  respectively. The DDS updates the next  $\hat{\phi}$  by adding the offset and prepares  $\gamma_1$  for the next phase detection.

In the process of the phase detector in Figure 5, for the  $n$ th input  $r_1(nT_s)$ ,  $r_1\gamma_1 = \sin(\phi - \hat{\phi}) - \sin(4\pi fT_s n + \phi + \hat{\phi})$ . Here we denote  $T_s$  as sampling period of received signal. As  $\phi_e = \text{LPF}(r_1\gamma_1)$  where LPF is the low pass filter, the high frequency component of  $r_1\gamma_1$  is eliminated and  $\phi_e \approx \sin(\phi - \hat{\phi})$ . If the phase is locked (*i.e.*,  $\hat{\phi}$  is close to  $\phi$ ),  $\sin(\phi - \hat{\phi}) \approx \phi - \hat{\phi}$ . Hence  $\phi_e \approx \phi - \hat{\phi}$ .

In Figure 5, the Loop Filter is the key part of PLL. There have been many proposals on design of loop filter [1], and the type and parameter of Loop Filter should be carefully chosen for different purposes. Here we adopt a second-order filter, *i.e.*, the proportional-plus-integrator [16] filter, as the Loop filter. It uses two updated variables  $\gamma_2, \gamma_3$  and two constant parameters  $k_1, k_2$ . Particularly, if  $k_2 = 0$ , it degrades to a first-order PLL.

We explain why the first-order PLL cannot be used in our case. When the phone is static and the PLL becomes stable after several cyclic loops,  $\phi_e \approx 0$  and  $\hat{\phi}$  is close to constant. Hence in Figure 5,  $\gamma_2 \approx 0$  and  $\gamma_2 + \gamma_3 \approx 0$  which infers  $\gamma_3 \approx 0$ . It means that  $k_2$  can be eliminated and the first-order PLL is sufficient. However, if the phone moves, and we still use the first-order PLL, the performance is good in case of high signal-to-noise ratio (SNR) but also limited by SNR. For instance, assume the user moves at a constant relative speed and  $\phi$  increases  $\Delta\phi$  per  $T_s$ , *i.e.*, the cyclical time of the loop. When the PLL becomes close to stable,  $\phi_e \approx \Delta\phi + \phi_s$  where  $\phi_s$  is error caused by random noises. If the SNR is high that  $\phi_s \ll \Delta\phi$ , we can set  $k_1 > \Delta\phi$  to let  $\hat{\phi}$  catch up with the variation of  $\phi$ . However, when increasing the value of  $k_1$ , as the magnitude of  $\gamma_2$  increases,  $\hat{\phi}$  becomes unstable and tends to be affected by noises. The bad case is that the phase, which is actually  $\phi$ , is intended to be locked or already locked to  $\phi + 2\pi$  or  $\phi - 2\pi$ . We call this phenomenon the *jitter* for convenience. The error of the corresponding displacement is  $\frac{v_a}{2\pi f} 2\pi \approx 1.8\text{cm}$  which affects the accuracy of position estimation in section 3.

To show the limitation in the experiment, we let the user hold the phone for a while, move the phone forward to the speaker and backward for three times, and finally stop at the starting point. In Figure 4, we show the result of PLL with different parameters, when the acoustic signal is weak, *i.e.*,  $l = 32\text{m}$ . In Figure 4a, the  $k_1$  is large enough to catch up with the real displacement. However,

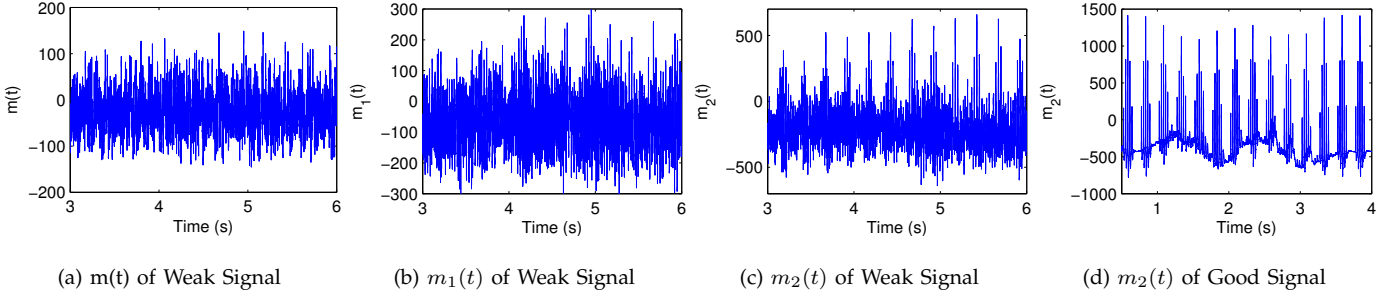


Fig. 6: Detection of the arrival time of the pulse.

it is affected by the noises and sometimes cannot lock when moving. It results in occasional jitters on the up-and-down curve. Then, the calculated displacement from the start to the end, which should be close to 0, accumulates to 17cm after total moving length of about 100cm. On the contrary in Figure 4b where  $k_1$  is small, the calculated phase displacement cannot catch up with the real phase and jitters frequently when moving. Hence, there is limitation of using first-order PLL for supporting both high-speeding moving and high noises.

Therefore, for solving the above problem, the updated component  $\gamma_3$  is added, which turns the first-order PLL into the second-order one.  $\gamma_3$  can be seen as the phase variation  $\Delta\phi$  per  $T_s$ , which corresponds to the relative speed from the phone to the speaker. If the PLL becomes stable, in each cyclic loop, the loop filter predicts next phase with the added  $\gamma_3$ , which results in  $\phi_e \approx \phi_s$ , instead of  $\phi_e \approx \Delta\phi + \phi_s$ . It means that it is no longer needed to set large  $k_1$  to let  $\hat{\phi}$  catch up with the dynamic  $\phi$ . Hence,  $k_1$  can be much smaller that the PLL is more robust to the noises. In Figure 4c, we choose the second-order PLL by setting  $k_2 \neq 0$ . Meanwhile,  $k_1$  is much smaller than the one in Figure 4a that the PLL is more robust to noises and does not cause observable jitters.  $k_1$  equals to the one in Figure 4b, but has no problem of catching up with the fast displacement for  $k_2 \neq 0$ . The accumulate displacement error is less than 2cm which is about at least 9 times more accurate than the one in Figure 4a.

## 5 POSITIONING BY SYNCHRONIZATION

Though we synthesize all the walking segments when user walks and turns, the problem is that the method has accumulated errors when we estimate the latter position by using the previous position, estimated walking directions and walking steps. Especially when the user is far away and loses the signal from the speaker for a long time, the error increases and the historical measured position can no longer be used. To solve this problem, we propose a synchronization mechanism that we leverage historical measurement to improve the robustness of WalkieLokie.

In synchronization, we additionally encode periodical pulses  $s_2(t)$  in sending signal and propose the demodulating method to detect the receiving time of the pulses. Since the pulses are periodical, the sending time of latter

pulses can be predicted, if we can accurately estimate the sending time of one periodical pulse. Hence, by using samples which can be directly used to calculate accurate position, we obtain the estimated distance, which infers traveling time  $t_l$  from the speaker to the phone. Then, we detect the receiving time  $\tau'$  of one pulse in these samples and get the accurate sending time of the pulse  $\tau = \tau' - t_l$ . Furthermore, the sending time of latter  $i$ th pulse equals to  $\tau_i = \tau + iT$ , where  $T$  is denoted as period of pulses. Hence, on obtaining the receiving time of  $i$ th pulse  $\tau'_i$ , we finally obtain real-time distance by using  $\tau_i$  and  $\tau'_i$  instead of the estimation method in Section 3.

### 5.1 Pulse Modulation

To design synchronization pulses  $s_2(t)$  and the detection algorithm, several problems should be addressed:

- Each speaker should not take much acoustic bandwidth in order to support more speakers in the room. Hence,  $s_1(t)$  and  $s_2(t)$  should be at the same frequency band, otherwise additional bandwidth for  $s_2(t)$  is needed. Moreover, bandwidth of  $s_2(t)$  needs to be narrow. However, it is challenging that  $s_2(t)$  should occupy more bandwidth if it can be successfully detected.
- $s_2(t)$  can also be used for displacement tracking by PLL. Otherwise, PLL will lose phase locks when processing  $s_2(t)$ .

Based on these requirements, we design  $s_2(t)$ :

$$s_2(t) = \begin{cases} \cos(2\pi ft + \pi \sin \frac{\pi(t-\tau_i)}{T_p}) & \tau_i \leq t \leq \tau_i + T_p \\ \cos(2\pi ft) & \text{otherwise} \end{cases} \quad (10)$$

where we construct pulses starting at  $\tau_1, \dots, \tau_i$ , and the duration of each pulse is  $T_p$ .

More specifically, Figure 7 shows an example of detected pulse when the user moves the phone forward and backward twice and then stops. We encode three adjacent pulses per  $T_2 = 0.25s$ . Three adjacent pulses can be seen as a compensated periodical pulse with the period  $T = T_2 = 0.25s$ . The time difference of the adjacent pulses is  $T_3 = 0.03s$ . In Figure 7a, the estimated displacement is smooth and have no jitters whenever the phone is static or moving. We zoom in the calculated phase to show the performance of PLL when there are pulses in  $s_2(t)$ : the calculated phase is not locked to the

real phase; instead, it seems that PLL has not detected the pulses that the phase is very smooth. Specifically, while the maximum variation of the real phase is  $\pi$ , the corresponding variation computed by PLL is less than  $0.4\text{rad}$ , which corresponds to the displacement of about  $1\text{mm}$ . The cause of the phenomenon is that the parameters  $(k_1, k_2)$  of PLL are very small, and does track the fast changing phase. Moreover, as the phase at the beginning of a pulse equals to the one at the end and the variation by PLL is small, the tracked phase finally becomes stable and the phase is the same with the one at the beginning.

### 5.1.1 Proof on Properties of Modulated Pulse

We prove that the pulse  $s_2$  does not take much acoustic bandwidth; and has little effects on the result of displacement tracking by PLL.

First, the central frequency of  $s_2$  is the same as the one of  $s_1$ , except that the phase changes when there is a pulse. Hence,  $s_1$  and  $s_2$  share the same frequency band. Second, since the bandwidth of the pulse is about  $\frac{\pi}{T_p}$  [17], we set  $T_p = 0.007\text{s}$  so that the bandwidth is about  $460\text{Hz}$ . As the minimum frequency is  $17000\text{Hz}$  when the acoustic is non-audible, and the maximum frequency which is supported by the phone is  $24000\text{Hz}$ , the maximum concurrent signals that WalkieLokie supports in one place is  $(24000 - 17000)/460 \approx 15$ . Actually, if the pulse has more narrow bandwidth, WalkieLokie will support more concurrent signals, whereas the pulse becomes harder to be detected. How to modulate signals with more narrow bandwidth and demodulate the signal more accurately is left for future work. Third, the component  $s_3(t) = \pi \sin \frac{\pi(t-\tau_i)}{T_p}$  is the phase shift of the sine signal. Furthermore,  $s_3(t)$  starts and ends at the same value 0, and the maximum value of  $s_3$  is  $\pi$ . Hence, the displacement will not be affected by the pulse theoretically.

### 5.1.2 Discussions of Pulse Modulation

**Choosing Parameters:** There is a trade off on choosing the parameters  $T_p, T_1, T_2, T_3$ , we show the analysis on choosing the parameters as follows:

a)  $T_p$ : As the bandwidth of pulses equals to  $\frac{\pi}{T_p}$ , smaller  $T_p$  results in wider bandwidth requirement and less simultaneous signals in the same room. On the other hand, greater  $T_p$  results in less accuracy of displacement tracking. The reason is that the pulses are regarded as noises in displacement tracking.

b)  $T_1$ : Since there are 3 adjacent pulses in one compensated periodical pulse,  $T_1 = T_2 - 3T_3$ .

c)  $T_2$ : Recall that  $T_2 = T$  which is the period of compensated pulses. Smaller  $T_2$  will enhance the accuracy of measuring the receiving time of pulses for we have more pulses for matching. However, if we choose smaller  $T_2$ , we may face the ambiguity problem. Specifically, denote the receiving time of a pulse is  $t_r$  and the sending time of periodical pulses is  $t_s + kT_2$ . The calculated distance is  $v_a(t_r - t_s - kT_2)$ , where  $k$  is an undetermined integer which also makes the distance undetermined. To get  $k$ , we further leverage the maximum distance from speaker to anchor, denoted as  $l_m$ . Since  $v_a(t_r - t_s - kT_2) < l_m$ , to get the unique solution of distance,  $v_a T_2$  should be greater than  $l_m$ . In our paper, we assume that  $l_m = 85\text{m}$  which infers  $T_2 = 0.25\text{s}$ .

d)  $T_3$ :  $T_3$  has limitation on its minimum value. Firstly, to avoid overlaps of adjacent pulses,  $T_3 > T_p$ . Secondly, there also should be intervals between adjacent pulses. We zoom in Figure 7a and find that PLL needs time longer than the duration of pulses to lock the displacement to the real value after the pulse terminates. Hence, if adjacent pulses are too close, PLL may become very unstable. If  $T_3$  increases, for  $T_1 = T_2 - 3T_3 > 0$ ,  $T_2$  may also increase which also affects the performance of WalkieLokie.

**Reducing Signal Conflicts:** As explained earlier, due to

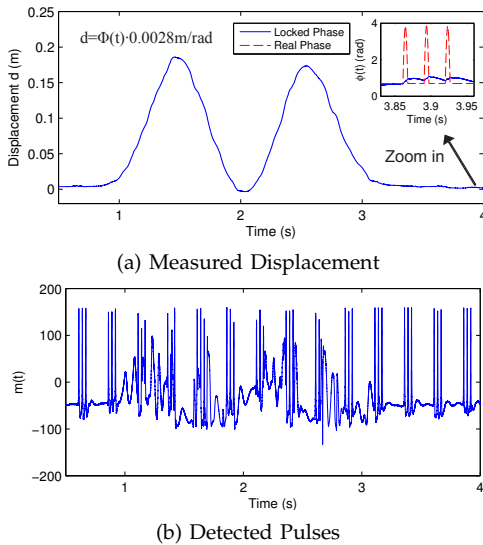


Fig. 7: Calculated displacement and pulses from the same signal.



the bandwidth limitation, our default parameter of pulse modulation supports 15 concurrent signals. Here, to reduce signal conflicts, we find that further optimizations can be made for different applications as follows:

a) Virtual business card sharing: In this case, users are usually close to each other, and we can choose to narrow the bandwidth of pulses in synchronization. Hence, WalkieLokie can support more users who broadcast signals simultaneously, while we only reduce the accuracy of pulse detection and long distance positioning, which are not much required.

b) Virtual shopping guide: We suggest that if there is requirement of more shopping guides, we can use only a few speakers for normal indoor localization, instead of just relative positioning. Our further evaluations in Section 6.4 prove that WalkieLokie supports unlimited number of shopping guides by simple and sparse deployment of speakers, *i.e.*, the smart device only receives signals from 2 speakers on average, but gains 1-meter accuracy.

## 5.2 Pulse Detection

We discuss how we detect the receiving time  $\tau'_i = \tau_i + t_l$  of the  $i$ th pulse by leveraging the component  $s_3(t)$ . Assuming the locked phase by PLL is  $\phi_r$  before the pulse starts, the expected pulse is  $\tilde{r}(t) = \cos(2\pi ft + \phi_r + \pi \sin \frac{\pi(t-\tau'_i)}{T_p})$ . Hence, for the received sample  $r(kT_s)$ , we compute the likelihood  $m(kT_s) = \sum_{i=k}^{k+T_p/T_s} r(iT_s)\tilde{r}(iT_s)$ , *i.e.*, when  $m(kT_s)$  reaches the maximum, the corresponding  $kT_s$  is the starting time of the received pulse. Note that, if we set expected pulse  $\hat{r}(t) = \cos(2\pi ft + \phi_r + \pi)$  and there is no pulse for the next  $T_p$  that  $r(t) = \cos(2\pi ft + \phi_r)$ ,  $\hat{m}(kT_s) = \sum_{i=k}^{k+T_p/T_s} r(iT_s)\hat{r}(iT_s)$  will reach the minimum. Actually,  $s_3(t)$  is the filtered version of pulse  $\hat{s}_3(t) = \pi$  that the pulse  $s_3(t)$  has narrower bandwidth. Accordingly,  $\tilde{r}(t) \approx \hat{r}(t)$  which means  $m(t)$  will reach the value close to minimum when there is no pulse in the next  $T_p$ . Hence, arrival time  $\tau'_i$  of the shape can be detected by  $m(t)$ .

### 5.2.1 Analysis on Design of Pulse Detection

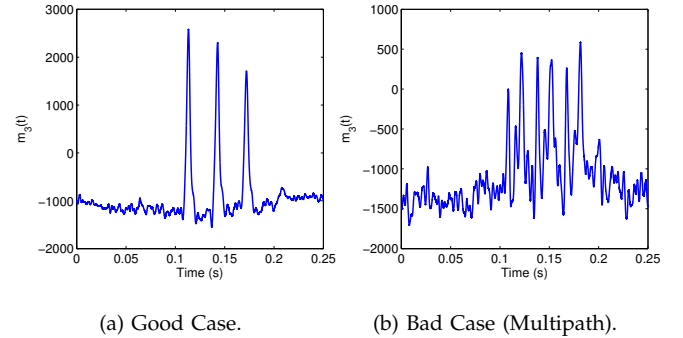
As mentioned earlier, our PLL takes  $s_2(t)$  as noises and only tracks  $s_1(t)$ . There are two advantages based on above results: 1) the pulses have very small effects on the tracked displacement. 2) For the variation is very small and the variation of  $\phi_r$  is stable when there are pulses, peaks of  $m(t)$  become clear to be detected. In Figure 7b,  $m(t)$  reaches the peak value (*i.e.*, 150), when there is a pulse at  $t$  and the bottom value (*i.e.*, -50) when there are almost no pulses. As a whole, it shows an interesting result that on demodulating  $s(t)$ , the peak of  $m(t)$  is very clear for synchronization in Figure 7b, while the corresponding calculated phase is very smooth for displacement tracking in Figure 7a.

We can also find that when the phone is static, the peaks corresponding to the pulses are clear. However,

they are unclear when the phone is moving. Furthermore, when the signal is weak, the periodical peaks cannot be detected by  $m(t)$  in Figure 6a due to noises. Hence, we make further solution to make the peaks more clear in case that the phone moves or the signal is weak. The solution is based on the observation that expected peaks still appear at expected time, though they sink in the noises. Meanwhile, random peaks have fewer chances to appear periodically. Hence, we assign  $m_1(t) = m(t - T_3) + m(t) + m(t + T_3)$  in Figure 6b, where the peaks are more clear to be identified in  $m_1(t)$ . Then, we assign  $m_2(t) = m_1(t - T_2) + m_1(t) + m_1(t + T_2)$  in Figure 6c, where the peaks can be easily detected. Moreover, when the phone is moving and the corresponding phase is in Figure 7a, the peaks are also very clear in Figure 6d.

### 5.2.2 Dealing With Multipath Effects:

We also find that the result of synchronization is affected by multipath effects, especially when the smart device is static. Hence, we make further study and improvement on pulse detection.



**Fig. 8: Pulse detection in case of multipath effects.**

We find that when the phone is static, there is another property which can be leveraged: the distance from the smart device to the dummy speaker is constant. Hence, we can use  $m_3(kT_2) = \sum_{i \in \{x | x=k \bmod T_2\}} m(iT_2)$ , which sums all the  $m_3(t)$  of pulses and make the detected time of pulses more clear. The result of summed  $m_3(kT)$  is shown in Figure 8. In Figure 8a when there is no multipath effect, there are 3 pulses in a period  $T_2$ . However, in Figure 8b, which is gathered from the shopping mall, there are 9 pulses at least, which means there are 2 additional paths reflected from walls or other objects. In this case, all the 3 paths are the possible pulses directly received from the dummy speaker.

After recognizing the possible multipaths, we make further step to filter the direct path. Specifically, we use the result of PLL, which corresponds to the displacement. As displacement tracking is less affected by multipath effects, we compare with the result of PLL and pulse detection when a user walks from one position and stops at another one. In this case, denote that the displacement by PLL is  $d$ , and the receiving time of pulses are in the set  $T_a = \{t_{a1}, t_{a2}, t_{a3}, \dots\}$  and  $T_b = \{t_{b1}, t_{b2}, t_{b3}, \dots\}$  at the start point and end

point respectively. Hence we obtain the receiving time  $(t_a, t_b) = \arg \min_{t_a \in T_a, t_b \in T_b} |(t_a - t_b)v_a - d|$ .

### 5.3 Positioning after Synchronization

Assume sending time of the next pulse is  $t_s$ , which is the result of synthesizing and the detected receiving time is  $t_r$ . Then distance  $l = v_a(t_r - t_s)$  and the distance at the horizontal plane is  $L = \sqrt{l^2 - h^2}$ . For direction estimation, we first calculate  $x$  and  $y$  using newly obtained  $L$ , previous  $s$  and  $d_i$ . For example, on calculating  $\psi'_1$  in Figure 2, assume  $l_1$  is obtained from synthesizing. Since  $x = -l_1 \cos \psi_1$  and  $y = l_1 \sin \psi_1$ ,  $l'_i$  in Eq. (1) has the following form

$$l'_i = \sqrt{l_1^2 \sin^2 \psi_1 + (-l_1 \cos \psi_1 + (i-1)s)^2} \quad (11)$$

Hence,  $\psi_1$  is obtained by  $\arg \min_{\psi_1} \sum_{i=1}^n e_i^2$  where  $e_i$  is calculated by Eq. (2), (11). Then  $\cos \psi'_1 = \frac{l_1 \cos \psi_1}{\sqrt{l_1^2 - h^2}}$ .

## 6 PERFORMANCE EVALUATION

In this section, we perform system evaluation by using two types of speakers: Samsung Galaxy Note 2 and normal dummy speakers. The speaker merely broadcasts acoustic waves and does not perform communications. We mainly use Google Nexus 4 to receive the acoustic signals. We do not make any modifications to the phone or jailbreak the operation system, and all the components, such as BPF, AGC, PLL, are implemented by the software. We evaluate the performance in an empty room, an office and the shopping mall. The micro benchmarks are made for position estimation and synchronization. We then evaluate the total performance where all the components are used.

Note that in our system, we do not measure the walking direction in World Coordinate System (WCS). The main reason is that this measurement is not necessarily needed in our system and we only calculate the relative position between walkers and speaker. Furthermore, accurately measuring walking direction in WCS is still challenging [15], which is mainly caused by unpredictable errors when using compass, especially in indoor environment. Therefore, to evaluate the accuracy of our system, we build and rely on relative coordinate system (RCS), instead of classical World Coordinate System (WCS). In RCS, we set the direction of the piecewise linear segment as X axis, and starting point of the segment as origin point. For instance, assuming  $Y = \sqrt{y^2 - h^2}$  and  $X = -x$  in Figure 2,  $(X, Y)$  is the position of the speaker when the user starts walking.

### 6.1 Position Estimation

We evaluate position estimation in several types of cases, *i.e.*, different related positions from the phone to the speaker, number of walking steps, users, orientation of devices, device diversity and environments, which may affect accuracy of the estimation.

#### 6.1.1 Positions

We make evaluation in an empty room to evaluate the performance at different places. In this experiment, the speaker is placed at different locations, *i.e.*,  $X = 2, 4, 6, 8m$ , and  $Y = 2, 4, 6, 8m$ . We let the user walk for 9 ~ 10 steps with the walking lengths of about 6m. The relative height  $h$  is about 0.3m. For each location, the user holds the phone in hand and walks for 35 times to gather samples, *i.e.*, we get 560 samples in this micro benchmark. Note that by using other smart devices, such as smart glasses, or smart watches, the user can have more comfortable experience. Due to that we only use IMU sensors and microphone which are frequently used in smart devices, we use the smartphone as smart device in the experiment. Then, we calculate the relative position for evaluating the accuracy of calculated distance  $L = \sqrt{X^2 + Y^2}$  and direction  $\psi' = \arccos(X/L)$ . Note that since the user walks for only several steps and the walking distance is short, we only evaluate the accuracy of the initial position  $(X, Y)$ .

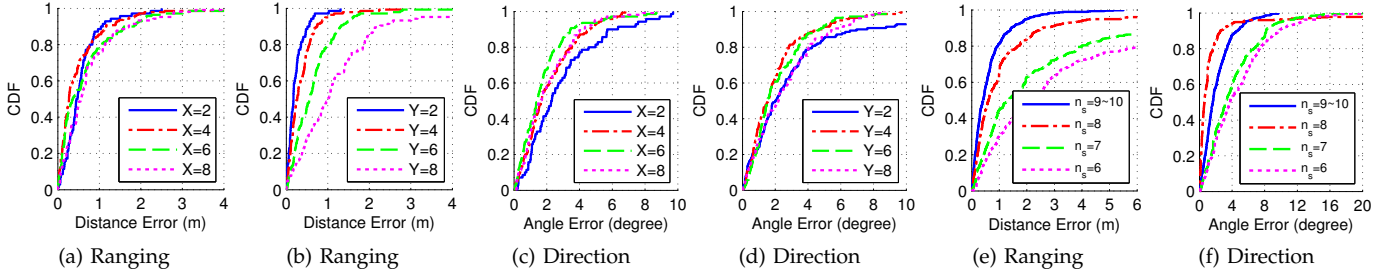
In Figure 9a, the accuracy of distance estimation is very close for different  $X$ . We further study the distribution of large errors in Figure 9b. We find an interesting fact that the errors are nearly proportional to  $Y$ . Hence, when  $Y = 2, 4, 6, 8m$ , the corresponding errors are within 0.35m, 0.55m, 0.97m, 1.88m at the percentage of 80%. The result is acceptable in our case for the user requires higher level of accuracy when s/he is close to the speaker. Furthermore, we use synchronization and synthesizing scheme achieves the accurate ranging in longer distances, instead of this position estimation scheme.

For direction estimation, it is still very accurate when the  $X$  or  $Y$  increases in Figure 9c, 9d. As a total, the mean of ranging and angle error is 0.63m and 2.46° respectively.

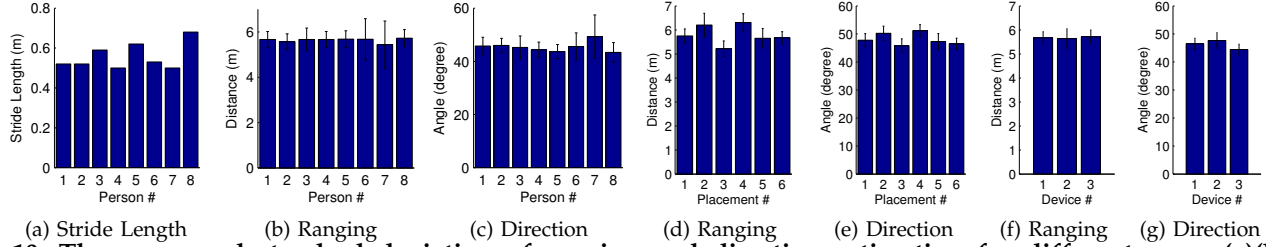
#### 6.1.2 Number of Steps

The accuracy of the position estimation depends on number of walking steps. We compare the results when the user walks for smaller number of steps  $n_s$  in Figure 9e, 9f. The samples are the same ones gathered in section 6.1.1, and the only difference is that we only use part of each sample which infers fewer walking steps. The results show that the ranging errors increase quickly when  $n_s$  reduces. The reasons are: 1) The user's stride length varies occasionally. 2) User's phone also shifts left and right regularly, *i.e.*, it does not move strictly in a line, when the user holds the phone and walks. As these facts will have less effects on the accuracy when  $n_s$  is larger, it can be foreseen that the accuracy will continue to be improved when  $n_s > 10$ , though it is already very accurate when  $n_s = 10$ .

The estimated direction is also affected by the smaller  $n_s$  in Figure 9f. But it is still acceptable that the angle errors are under 8° at the percentage of 80%, when  $n_s = 6$ . As a whole, when  $n_s$  is small, the accuracy is enough



**Fig. 9: The accuracy of ranging and direction finding 1) when the user starts walking at different positions (a)(b)(c)(d), 2) when the user walks for smaller number of steps (e)(f).**



**Fig. 10: The mean and standard deviation of ranging and direction estimation for different users (a)(b)(c), placements (d)(e), and smart devices (f)(g).**

for direction estimation of surrounding speaker, while it requires latter synthesizing scheme to obtain accurate distance.

### 6.1.3 Users

Different user has different stride length and user motion when the user walks, which causes variation of displacement patterns displacement  $d_i$  and might affect the positioning result. Hence, we recruit 8 volunteers in this experiment: each user walks in a line of about 6m for 35 times where  $(X, Y) = (4, 4)$ .

We have the following observations in Figure 10: The standard deviations (std) of the ranging and direction are small for most users. In Figure 10a, the person 1,2,4,6,7 have small stride lengths while the rest ones have bigger length, but the result is similar for all the users (except for the person 6,7).

The results infer that the stride length is very stable and the positioning accuracy is not much affected by variation of stride length, though the stride length between different users may be much different.

### 6.1.4 Orientation of Speaker and Microphone

We consider the cases when the speaker or the microphone faces to different directions: (1) (default) the microphone faces to the sky, and the speaker faces to the walking line. (2) microphone, facing to the front. (3) microphone, perpendicular to the walking direction and facing to the speaker. (4) microphone, facing to the ground. (5) microphone, perpendicular to the walking direction and speaker is at the back of the microphone. (6) speaker, facing to the ground. The result in Figure 10d, 10e shows that the std is small in all cases and the result is very stable.

We also find that the mean value of distance increases when the signal is weaker in case (2), (4) and

decreases when signal is stronger in case (3). The reason is that when the signal is weak, PLL will lose some signals and the displacement decreases, which makes the calculated distance become larger. Hence, based on our measurements in displacement tracking, we make calibrations on the calculated PLL. More specifically in case (1) that the displacement  $d = 1.22 \frac{v_a}{2\pi f} \Delta\phi$ ; if  $d > 0$ , and  $d = 1.69 \frac{v_a}{2\pi f} \Delta\phi$ , if  $d < 0$ , where  $\Delta\phi$  is the tracked phase shift. Note that we make calibration with constant factor (*i.e.*, 1.22), for the environment has limited effect on the result of PLL when the signal is strong enough. However, when  $d < 0$ , which means the speaker is at the back of the walking user,  $d$  is usually not used for position estimation if the tracked phase is abnormal (*e.g.*, when WalkieLokie cannot detect pulses from the phase).

### 6.1.5 Device Diversity

We test several Commercial Off-the-Shelf (COTS) smart devices as acoustic receivers: (1) Nexus 4, (2) Samsung Galaxy Note 2. (3) Nexus 7. We choose  $(X, Y) = (4, 4)$  as the start point of walking, and the error of position estimation is shown in Figure 10f, 10g. The result shows that these smart devices have similar performance.

We also use normal dummy speakers as acoustic speakers when we make experiment in a large shopping mall, for we consider the case that the normal speakers serve as virtual shopping guides.

**Calibration of clock drift:** We find some interesting phenomenon: different from the previous smart devices, the normal speaker has serious clock drift and needs to be calibrated. For instance, when a speaker is supposed to broadcast signal at 19000Hz, the actual received signal is 19007Hz. If the frequency drift is 0.1Hz, the error of distance measuring is about  $600 \times 340 \times 0.1 / 19000 = 1.07m$ , when the smart device performs synchronization for 10 minutes. To solve this problem, our design of PLL

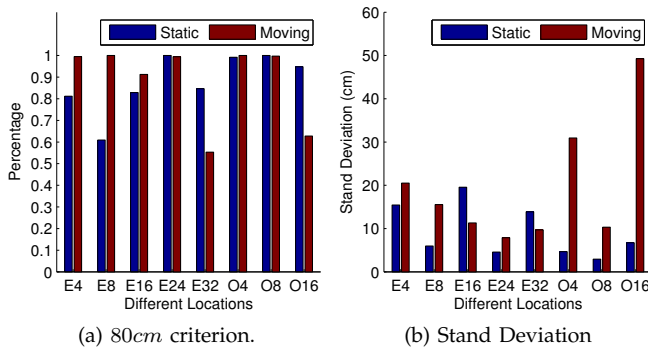
measures the precise clock offset when the receiver is static for only a few seconds. In this case,  $\gamma_2$  in Figure 5 rapidly converges to a constant value. As  $\gamma_2$  equals to the phase shift per sampling time  $T_s$ , the frequency offset equals to  $\frac{k_2}{2\pi T_s}$ . Hence, once we let the smart device be static for a few seconds, the precise frequency offset is obtained. Afterward, we calibrate the clock drift in real-time using the constant frequency offset and there is not any clock drift after calibration.

### 6.1.6 Environments

We compare the accuracy of position estimation in the empty room and at different locations in the office. We find that it shows the similar results. We further evaluate the effects in a shopping mall in the latter subsection.

## 6.2 Synchronization

In Figure 11, we choose 8 locations in an empty room and the office to evaluate the performance of synchronization. For example, E32 means that the experiment is in the room and the distance from the smart device to the speaker is 32m, and O16 means that it is in the office and the distance is 16m. In each position we test two cases: the phone is static or moving back and forth without stop. For each case, the phone records the audio for 100 seconds, which means there are 400 signals for synchronization in the samples. Then, we evaluate the accuracy of pulse detection. For easier understanding of our results, the error of arrival time  $t_e$  is converted to distance measurement error  $l_e = v_a t_e$ . For instance, if the error is the time interval of 1 acoustic sample, i.e.,  $t_e = \frac{1}{44100}s$ , the corresponding distance error is  $l_e \approx 0.8cm$ .



**Fig. 11: (a) Percentage of successful experiments at different locations (b) standard deviation.**

Since we find that there are occasional significant errors ( $> 3m$ ), we first set threshold  $l_t = 80cm$  and evaluate ratio of successful detection that  $l_e < l_t$ . In Figure 11a, the successful detection rate is above 80% for most cases when the phone is static. When the phone is moving, the performance is good as well if the distance is within 24m and 8m in the empty room and office respectively. In some cases the rate is close to 100%.

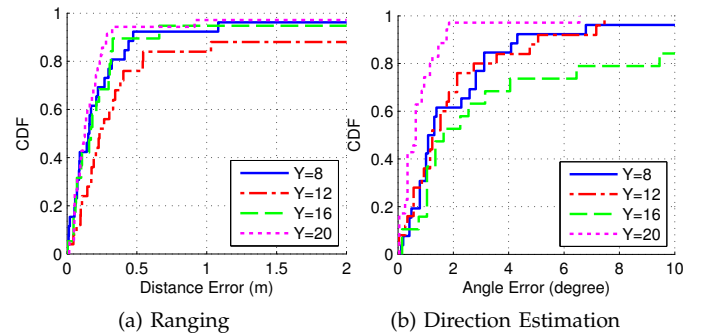
There is also an exception that at location E8 when the phone is static, the rate is only 61.0%, while it reaches 100% at the same place when the phone is moving. So,

we conduct the experiment again at the same place, and the result is close to the previous one. We suppose it is caused by the multipath effects: the phase  $\phi_r$  changes according to the mixed signals and becomes stable when it is static, which affects the result of pulse matching. The reason of high successful rate in case of moving phone is that: though it is also affected by multipath, the phases of reflected signal at different positions are irregular. In other words, the PLL locks the phase of the signal directly from the speaker, i.e., the multipath signal is regarded as noises by PLL. Hence, the performance is better when the phone is moving. We find the location E4, E8, E16 also have the same phenomenon, which validates our hypothesis. Actually, this is a good result for WalkieLokie: when the user is walking, the synchronization result is very good and can be directly used for synthesizing; when the user is walking, as the successful detection rate is above 60%, WalkieLokie collects enough samples and then determines the most possible receiving time. In Figure 11b, we show the standard deviation of results in case of successful detection. The std in most cases are around 10cm expect that the std is 30.9cm and 49.2cm when the phone is moving at O4 and O16 respectively.

## 6.3 Positioning after Synchronization

We evaluate the performance of WalkieLokie which uses position estimation and synchronization in the following steps:

- 1) The user walks in a line where the initial coordinate of the speaker is (4, 4). In this step, we calculate the distance through position estimation and then calculate the sending time of periodical signals  $s_2(t)$  by synchronization.
- 2) The user then turns, walks and stops at the position where relative coordinate of speaker is (X, Y).
- 3) The user walks again for about 6m. The position, which is supposed to be (X, Y), is then computed according to the sending time and the received samples in this short duration of walking.



**Fig. 12: Accuracy of positioning by synchronization.**

We conduct the experiment in the empty room and the office. Specifically, we set  $(X, Y) = (4, 12)$  and  $(4, 20)$  in the empty room to gather the samples and  $(4, 8)$  and  $(4, 16)$  in the office.



In Figure 12a, the ranging errors are under  $0.32m$  and  $0.66m$  at the percentage of 80% and 90% for most cases. It means that both position estimation and synchronization achieve considerable accuracy. There are also occasional errors for each cases which are greater than  $2m$ . It is caused by the multipath effects in synchronization. Especially for the case of  $Y = 12m$  in the empty room, the big errors are at the percentage of 12%. We can find the corresponding results at E8 and E16 shown in Figure 11a, where the successful detection rate is also much lower than other cases in synchronization. Actually, since the successful detection rate in synchronization is above 80% for most cases, the result would converge to the correct value if given enough time and the abnormal result would be eliminated. Hence, we conclude that the ranging results are very good in these cases.

#### 6.4 Putting it All Together in a Severe Environment

We evaluate WalkieLokie in a shopping mall, where the environment is quite severe for acoustic based systems: the shopping mall itself is broadcasting loud audios; there are always people walking around who blocks the sight line of speakers or blocks the road that we have to turn walking direction. Furthermore, as it may affect the business if we set up speakers on the ceiling and conduct frequent debugging (which may have better results), we only put the speakers at the side of the aisles, as shown in Figure 13, 14a. Hence, our system has to deal with serious NLoS effects.

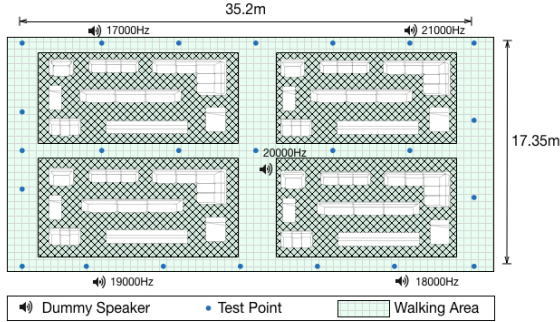


Fig. 13: Map of the shopping mall.

We evaluate the performance of positioning in two cases: a) *relative positioning by one speaker*. b) *absolute positioning by 5 speakers* (like normal indoor localization). We choose a  $35m \times 17m$  area (about  $600m^2$ ) in Figure 13, and put 5 normal dummy speakers in this area. Each speaker broadcasts signals at different central frequency, which are inaudible and not discovered by surrounding customers. We emulate the behavior of normal shopping users in evaluation: the experimenter stands at a test point and walks for a few steps (less than  $6m$ ) in a line; then he stops or turns the direction and continues walking, and so on. We gather 8 samples per point. Hence, we can evaluate the performance when leveraging all the walking segments to get the position.

We set central frequency of the speakers to 17000Hz, 18000Hz, 19000Hz, 20000Hz, 21000Hz, respectively. The

smart device differentiates the signals by using the sub-component BPF in the Figure 1b. For example, if we need to analyze the signal of the second speaker (18000Hz), we set the frequency band of BPF which filters the signal at 18000Hz, and other signals are blocked.

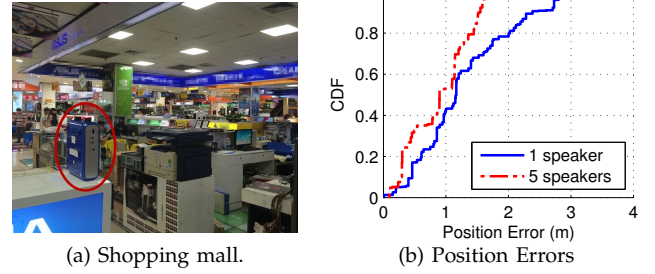


Fig. 14: (a) Shopping mall and the dummy speaker. (b) Result of relative positioning (using 1 speaker), and absolute positioning (using 5 speakers).

The results show that these 5 speakers have much different performances in relative positioning, though they are the same product model. The signal of speaker at 17000Hz only covers 13% of the area, but the signal of speaker at 19000 and 20000 covers about 54% and 51% of the area. The reason of this diversity may be caused by several facts: anchor positions, quality of different anchor speakers, etc. We leave the study on configuration of speakers in our future work. Totally, the average coverage per speaker is 38%, which is about  $222m^2$  in our specific area.

We show the relative position errors when using one speaker in Figure 14b. Note that we only calculate the accuracy of the relative position where the starting point is covered by the signal of the speaker. Though we can still estimate position according to historical positioning result when there is no signal, we exclude the results of this case and obtains the direct result. The results show that for one speaker, the position errors are under  $1.2m$ ,  $2m$  at the percentage of 50% and 80%. The mean error of relative positioning is  $1.28m$ .

We also explore the localization capabilities when all 5 speakers are used as anchors. We evaluate the errors at all points and the results show that the position errors are under  $1.5m$  at the percentage of 90%. Since the average coverage per speaker is 38%, the smart device can receive audio from  $38\% \times 5 \approx 2$  speakers on average. The accuracy is intuitively better when using multiple signals for localization.

#### 6.5 Overhead

The computation overhead is mainly caused by 3 components: displacement tracking (Including BPF, AGC, PLL), pulse detection and position estimation. We run WalkieLokie using matlab R2013a on Mac OS, and the CPU is 3.1GHz Intel Core i5. For 1 second of received samples, phase tracking, pulse detection, and position estimation takes 0.09s, 0.12s, 0.05s respectively. In fact, there is a trade-off between the overhead and accuracy. For example, we can use infinite BPF instead of finite



BPF, which reduces the computation overhead significantly, but incurs larger errors. For the smart devices, it is recommended to send the recorded samples to cloud server, and obtains the result from the cloud, which requires much less computation overhead, meanwhile with low energy consumption.

## 7 RELATED WORK

### 7.1 Ranging

There have been many localization systems which are based on ranging [2], [7], [8], [13]. They achieve considerable accuracy of ranging, but require special hardwares for synchronization purpose. Specifically, the sender records sending time of signal which is used for ranging, while the receiver detects the arrival time of the signal. Each individuals calculate the sending time or arrival time independently without referring any time information on other devices. Hence, synchronization among devices is needed. In Bat System [2], the base-station uses radio channel and communications for synchronization. Cricket [13] uses special device to send the RF signal together with the ultrasound signal at the same time. Then the receiver obtains the distance according to the different traveling time of the two signals. Guoguo [7] uses RF signals to synchronize all the acoustic anchors, the location can be obtained according to the differences of the receiving time by the phone. BeepBeep [12] calculates the distance between the phones. It solves the synchronization problem by letting two phones emit acoustic signals and exchange the sending and receiving time via wireless channel.

WalkieLokie uses dummy speaker to implement synchronization and ranging. The synchronization information is obtained by a novel position estimation method that it does not need any special hardwares or additional communication channels. The other difference is that these systems are only based on ranging results of anchors which requires multiple speakers ( $\geq 3$ ), while WalkieLokie also implements direction estimation from phone to speaker and only one speaker is needed for localization.

### 7.2 Direction Estimation

Most methods on direction estimation also require specialized hardwares, which use the directional antenna [5], [10], [18] or the antenna array [4], [18], [19]. For example, by rotating the beam of directional antenna, a receiver can pinpoint the direction of the AP as the direction that provides the highest received strength [18]. For the antenna array [4], [18], [19], the receiving time of the signal by each antenna is different, and magnitude of the difference corresponds to angle of the arrival signal.

There have been proposals without requirement of specialized hardwares as well. [21] emulates the functionality of a directional antenna by rotating the phone around the user's body, to locate outdoor APs. [14]

leverages multiple microphones of the smartphone and communication channels for positioning within 4 meters, which is used for short-distance positioning and phone-to-phone games. Some other methods leverage Doppler effects by swinging [11] or shaking [3] the phone. [20] calculates direction by head nodding or shaking using smart glasses. They are based on different frequency shift when the phone are moving at different directions. Compares to [3], [11], WalkieLokie makes further steps that a user can obtain direction without any additional actions on the phone so that s/he can get the real-time direction while walking. Furthermore, [3] requires only the speakers as anchors as well, but does not address the ranging problem, while WalkieLokie can compute both the direction and distance from the phone to the speaker.

## 8 CONCLUSION

We propose and implement WalkieLokie, a localization scheme that calculates the relative position from a smart device to a dummy speaker. The dummy speaker only needs to emit acoustic signals at non-audible frequency, so that COTS speakers can serve as anchors. Furthermore, WalkieLokie directly obtains both distance and direction from smart device to speaker, which is quite different from existing localization systems that are capable of obtaining only the distance or the direction. As a result, WalkieLokie only requires one anchor for localization, while others need multiple anchors for calculating the final position, such as trilateration. By pushing the limit of the anchor's number, WalkieLokie is not only capable of indoor localization, but also has a potential for wider applications, such as augmented-reality or mobile social applications.

## REFERENCES

- [1] BEST, R. *Phase Locked Loops: Design, Simulation, and Applications*. Professional Engineering. McGraw-hill, 2007.
- [2] HARTER, A., HOPPER, A., STEGGLES, P., WARD, A., AND WEBSTER, P. The anatomy of a context-aware application. In *MOBI-COM* (1999), pp. 187–197.
- [3] HUANG, W., XIONG, Y., LI, X.-Y., LIN, H., MAO, X., YANG, P., AND LIU, Y. Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones. In *INFOCOM* (2014).
- [4] JOSHI, K., HONG, S., AND KATTI, S. Pinpoint: localizing interfering radios. In *NSDI* (2013), pp. 241–253.
- [5] KIM, M., AND CHONG, N. Y. Direction sensing rfid reader for mobile robot navigation. *Automation Science and Engineering, IEEE Transactions on* (2009), 44–54.
- [6] LIU, H., GAN, Y., YANG, J., SIDHOM, S., WANG, Y., CHEN, Y., AND YE, F. Push the limit of wifi based localization for smartphones. In *MobiCom* (2012).
- [7] LIU, K., LIU, X., AND LI, X. Guoguo: enabling fine-grained indoor localization via smartphone. In *MobiSys* (2013), pp. 235–248.
- [8] LIU, K., LIU, X., XIE, L., AND LI, X. Towards accurate acoustic localization on a smartphone. In *INFOCOM* (2013), pp. 495–499.
- [9] MADSEN, K., NIELSEN, H. B., TINGLEFF, O., AND MODELLING, M. Imm methods for non-linear least squares problems. Technical University of Denmark, DTU, 2004.
- [10] NICULESCU, D., AND NATH, B. Vor base stations for indoor 802.11 positioning. In *MobiCom* (2004), pp. 58–69.
- [11] NISHIMURA, Y., IMAI, N., AND YOSHIHARA, K. A proposal on direction estimation between devices using acoustic waves. In *MobiQuitous* (2012), pp. 25–36.

- [12] PENG, C., SHEN, G., ZHANG, Y., LI, Y., AND TAN, K. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *SenSys* (2007), pp. 1–14.
- [13] PRIYANTHA, N. B., CHAKRABORTY, A., AND BALAKRISHNAN, H. The cricket location-support system. In *MobiCom* (2000), pp. 32–43.
- [14] QIU, J., CHU, D., MENG, X., AND MOSCIBRODA, T. On the feasibility of real-time phone-to-phone 3d localization. In *SenSys* (2011), pp. 190–203.
- [15] RAI, A., CHINTALAPUDI, K. K., PADMANABHAN, V. N., AND SEN, R. Zee: zero-effort crowdsourcing for indoor localization. In *MobiCom* (2012), pp. 293–304.
- [16] RICE, M. *Digital Communications: A Discrete-Time Approach*. Prentice Hall, 2008.
- [17] SAHU, O. P., AND GUPTA, A. K. Measurement of distance and medium velocity using frequency-modulated sound/ultrasound. *IEEE T. Instrumentation and Measurement* 57, 4 (2008), 838–842.
- [18] SUBRAMANIAN, A., DESHPANDE, P., GAOJGAO, J., AND DAS, S. Drive-by localization of roadside wifi networks. In *INFOCOM* (2008), pp. 718–725.
- [19] XIONG, J., AND JAMIESON, K. Arraytrack: A fine-grained indoor location system. In *NSDI* (2013), pp. 71–84.
- [20] ZHANG, L., LI, X., HUANG, W., LIU, K., ZONG, S., JIAN, X., FENG, P., JUNG, T., AND LIU, Y. It starts with igaze: visual attention driven networking with smart glasses. In *MobiCom* (Maui, HI, USA, 2014), pp. 91–102.
- [21] ZHANG, Z., ZHOU, X., ZHANG, W., ZHANG, Y., WANG, G., ZHAO, B. Y., AND ZHENG, H. I am the antenna: accurate outdoor ap location using smartphones. In *MobiCom* (2011), pp. 109–120.